CIE Independent Report
SEDAR Procedural Workshop on Uncertainty
Charlotte, North Carolina. 22-26 February 2010

Dr. Noel Cadigan
Fisheries and Oceans Canada,
Northwest Atlantic Fisheries Center
St. John's, NL, Canada

Executive Summary

The workshop report was not available so a summary of findings and recommendations from the workshop are not yet available.

The workshop agenda was ambitious. The theory and software technology for representing uncertainty in stock assessment is still evolving which made it difficult to recommend best practices. The workshop did not make specific recommendations for expressing confidence in estimated parameters. Instead, a range of options were discussed. Also, a range of options was presented for accounting for inter- and intra-model uncertainty.

Past approaches to characterizing and presenting uncertainty in stock assessments and projection analyses conducted under the SEDAR process were reviewed. The approaches were somewhat ad hoc, reflecting the different data available for assessments and the evolving nature of practical methods for quantifying uncertainty. This situation seemed to be basically the same in reviews of approaches used in other regions.

An important source of uncertainty identified was the specification of the stock assessment model. Two approaches were proposed to account for this: the empirical approach, and model averaging. Both approaches have difficulties yet to be resolved and neither could be recommended as currently most appropriate.

Background

SEDAR procedural workshops provide an opportunity for focused discussion and deliberation on topics that arise in multiple assessments and are structured to develop best practices for addressing common issues across assessments. The fourth procedural workshop considered methods to address uncertainty in SEDAR assessments. Topics included: appropriate confidence intervals in both parameter estimates and projection outputs; characterizing and expressing assessment uncertainty beyond that reflected in confidence intervals; sensitivity analyses and recommendations on standard sensitivities; uncertainty and overall risk evaluation and especially the risk of overfishing occurring.

The CIE reviewer was tasked with conducting an independent peer review of NMFS science in compliance the Statement of Work (SoW) and predetermined ToRs (attached in Appendix 2) for the SEDAR Procedural Workshop on Uncertainty.

Role of reviewer

I attended the SEDAR Procedural Workshop on Uncertainty held in Charlotte, North Carolina during 22-26 February 2010. I reviewed presentations and reports and participated in the discussion of these documents, in accordance with the SoW and ToRs (see Appendix 2). The review is structured according to the required format and content described in Annex 1 of Appendix 2.

There were no background documents to review.

Subsequent to the meeting I reviewed some literature related to accounting for uncertainty in fish stock assessments, and I provided a summary of this literature (**Appendix 4**) as well as my own views in this CIE report.

Summary of findings

The workshop report was not available. The following summaries are my own, and not consensus views of the workshop.

ToR 1: Review and discuss past approaches to characterizing and presenting uncertainty in stock assessments and projection analyses conducted under the SEDAR process.

The following presentations related to ToR 1 were reviewed. I provide the abstracts to help motivate my comments, which are provided at the end of this section.

<u>South Atlantic assessment uncertainty methods. E. H. Williams, SEFSC, Beaufort Laboratory</u>

Stock assessments for South Atlantic Fishery Management Council (SAFMC) managed fish species have been primarily conducted by staff located at the Southeast Fisheries Science Center (SEFSC), Beaufort Laboratory in North Carolina. The species assessed by the Beaufort staff through the SEDAR (Southeast Data, Assessment and Review) process are primarily composed of members of the snapper-grouper complex. Stock assessments for ASMFC managed Atlantic menhaden are also conducted by staff at the Beaufort Laboratory.

A review of the data used for the South Atlantic stock assessments indicates many areas of uncertainty including ageing error, age sampling, abundance indices, recreational landings data, historic landings data, and discard data. In some cases there are important sources of information for which the data are missing. These missing data include fishery-independent indices, shrimp trawl bycatch, spatial/depth data, and environmental linkages. Missing data and uncertainties in the data in large part drive how uncertainty is handled and ultimately expressed in the South Atlantic.

Three types of model uncertainty were considered for the South Atlantic stock assessments: structural, parameter, and projection uncertainty. These sources have been managed in different ways. Structural uncertainty has been addressed in the South Atlantic by applying different stock assessment models and software packages. The list of model types used in the South Atlantic includes forward-projecting, statistical catch-at-age models, surplus-production models, stochastic stock reduction analysis, and stock synthesis 3. Typically a set of two or three of these models are applied to the same species for comparison, however the output of these multiple models is not combined to encompass among model uncertainty.

Parameter and output uncertainty has been characterized in South Atlantic stock assessments through sensitivity runs, inverse Hessian estimates, data bootstrapping, and Monte Carlo bootstrapping. Sensitivity runs have been used primarily to illustrate model responses to various perturbations of input and model structure, but never combined to produce comprehensive uncertainty estimates. Uncertainty estimates derived from the inverse Hessian matrix tend to be underestimates of true uncertainty because often key parameters are fixed and the model typically includes penalty functions in the total maximum likelihood estimates. Both data and Monte Carlo bootstrapping tend to be more comprehensive in characterizing total uncertainty in model output.

Uncertainty in projection analyses in the South Atlantic have been accomplished using either simple bootstrap methods, in which only recruitment residuals are re-sampled, or as part of a full Monte Carlo bootstrap which carries uncertainty from the model fitting process forward in to the projections. An important property to consider with any projection analysis is benchmark consistency by ensuring that populations reach a long-term equilibrium consistent with the fishing rate being applied (i.e. fishing at FMSY achieves BMSY long term). Conditioning projections to conform to probabilities of overfishing (P*) is an important part of the management advice process. The SAFMC acceptable biological catch (ABC) control rule relies on projections for various P* levels.

It is noted that the framework created by the re-authorized Magnuson-Stevens Act has in effect placed a monetary value on uncertainty. This arises under prescriptive ABC control rules which link reductions in ABC to levels of uncertainty. This linkage will likely result in external pressures to handle uncertainty in a much more rigorous, consistent ways compared to past practices.

Bayesian uncertainty techniques as applied to sharks. K. Andrews, SEFSC Panama City Laboratory

Shark stock assessments are data-limited, both in the information about their life histories, but also in the amount of catch and index values available. Our presentation reviews the biological modeling of life history parameters and the evolution of modeling that has occurred for Gulf of Mexico shark stocks. Although sophisticated modeling techniques are available, they require large amounts of data that are often unavailable or highly uncertain for shark stocks. In addition, the modeling of shark bycatch in the shrimp trawl fishery, which is often a substantial portion of the catch for some shark species, is in a state of flux. The modeling technique is being revised to incorporate the effects Turtle Excluder Devices (TEDs) have had on the total number of sharks taken. Finally, catch rates are derived from both fishery-dependent and –independent surveys. The independent surveys available for sharks are often too limited in space or time to be representative and the dependent surveys are subject to high levels of criticism and uncertainty. Overall, the shark assessments in the Gulf of Mexico present uncertainty in a Bayesian framework. We assign priors on important life history parameters, such as pup survivorship and steepness, based on data or expert opinion in order to characterize our uncertainty. The model outputs include either likelihood profiles or MCMC runs to create distributions, rather than point estimates for parameters of interest.

Gulf of Mexico assessment uncertainty methods. S. L. Cass-Calay, SEFSC, Miami.

The objective of this presentation is to introduce the methods used to characterize scientific uncertainty during past and current SEDAR stock assessments in the Gulf of Mexico (GOM). Species considered in this presentation include: vermilion snapper, greater amberjack, gray triggerfish, king mackerel, red grouper, gag grouper and red snapper. In general, uncertainties examined during GOM SEDAR assessments include: uncertainty in data inputs (e.g. discards and bycatch, release mortality, catch statistics) and parameter uncertainty (e.g. natural mortality, steepness, selectivity and catchability). In GOM assessments, uncertainty is generally explored using sensitivity runs and/or bootstrapping. A suite of sensitivity runs are often chosen to represent plausible “states of nature” and presented to the Gulf of Mexico Fisheries Management Council - Science and Statistical Committee (GMFMC SSC) for consideration. To date, a single “base” model has been adopted by the SSC in order to develop management advice. For a given base run, complete characterization of model uncertainty is often not carried forward into the projections used for management advice. Instead, the projected stock status, yield, etc. are generally bootstrapped using only the index residuals and/or recruitment variance. To date, no attempt has been made to carry structural uncertainty (e.g. sensitivity runs) into the determination of OFL/ABC/ACL/ACT using model weighting techniques, although the SSC has discussed the merit of this approach. To date, not all GOM assessments use techniques that are ideal, or even appropriate for the methods outlined by the GMFMC SSC to determine OFL/ABC/ACL/ACT levels. However, future assessments are expected to be compliant with those objectives.

<u>Beaufort approach. P. B. Conn, SEFC, Beaufort.</u>

In this talk, I review several methods used by NMFS-Beaufort assessment scientists in recent SEDAR assessments and outline approaches under consideration for future assessments. Throughout the talk, I concentrate on uncertainty conditional on a given model structure and dataset (thus, additional uncertainty attributable to differences between models and to alternative data streams are beyond the scope of this presentation). Uncertainty in previous SEDAR assessments has primarily been addressed using bootstrap-based approaches. In particular, most assessments have used bootstrapping of spawner-recruit residuals to account for uncertainty in management benchmarks (e.g., FMSY, BMSY). Alternatively, several assessments (e.g., snowy grouper, red grouper) have incorporated the Monte Carlo bootstrap, which combines a data bootstrap with a Monte Carlo procedure that accounts for uncertainty in parameters that are modeled as fixed parameters within stock assessments (e.g., natural mortality, discard mortality). The latter procedure assigns a prior distribution to fixed parameters; variation in model runs with 'fixed' values sampled from their prior distributions results in additional uncertainty associated with these parameters. Using a recent assessment model (SEDAR 19 red grouper), these approaches are contrasted with other possible approaches for accounting for uncertainty, including asymptotic, Hessian based methods, as well as Bayesian approaches (including maximum a posterior [MAP] estimation). This comparison illustrates how inclusion of additional sources of error can drastically impact the variance associated with distributions of management benchmarks. Finally, I describe how a recently developed method ('inverse prediction') can be used to estimate uncertainty distributions when a simulation study is used to relate estimated assessment parameters to those used to generate data. In this case, the assessment model is regarded as a 'black box' and the relationship between true and estimated quantities is estimated empirically. When a real life assessment with the same structure and data sources is used to estimate parameters of interest, the results of this experiment can be used to calibrate estimated values to true, unknown values (thus accounting for bias), and to also estimate uncertainty about these parameters.

Reviewer's views

The review part of this ToR was addressed well. I think more attention could have been given to inadequacies in the past approaches to characterizing uncertainty that the workshop could attempt to fix or provide guidance on.

Overall, I found that rather ad hoc approaches have been used to characterize uncertainty about stock status and projections under the SEDAR process. This is typical of most stock assessments, at least those that seriously attempt to account for uncertainty. There are many stock assessments that do not. The ad hoc nature of approaches to account for uncertainty is not a criticism but it simply reflects the evolving nature of the theory and technology available. Best practice has evolved fairly rapidly in this area, and it will likely to continue to evolve for some time yet. Different assessment methods are needed for different types of stocks and stock data, and this may also affect the methods used to quantify uncertainty.

More objective sensitivity analyses could also be informative, although high-dimensional sensitivity analyses can be computationally prohibitive. Cadigan and Farrell (2002) presented an approach that is computationally more tractable, but it is only useful for perturbing single model components like M for each age and year, or catch, etc.

The following comments were motivated by the presentations and so I present them in that format. However, many of the comments are fairly general.

<u>South Atlantic assessment uncertainty methods</u>

Sensitivity runs are used to characterize structural (i.e. model) and parameter (e.g. M) uncertainty. My concern is that sensitivity runs really address robustness and not uncertainty. The range of results from sensitivity analyses does not provide a measure of uncertainty that I can interpret. I am not sure if wide is bad, or narrow is good. A wide range may result from including in the sensitivity analyses unrealistic models that fit the data poorly. A narrow range may occur if the assessment/review team somehow missed the major "axis" of uncertainty. In either case the range has not provided a measure of uncertainty that would be relevant or useful for advice. Usually one cannot say that there is a low probability that stock status is outside the sensitivity interval.

Missing data, including missing catch statistics, were identified as important sources of uncertainty. This is a common problem. Bousquet et al. (2010) investigated a censored estimation approach to address missing catch. Censoring refers to a technique commonly used in survival analysis and industrial reliability. A censored observation is one for which we do not know the exact value but about which we have partial information, such as a lower bound (right-censoring) or an interval known to cover the exact value (interval-censoring). This may be a useful way to incorporate uncertainty about missing catch, but further research is required. If the missing catch is a large fraction of the total then it may be just as well to consider the catch as unknown, and pursue catch-free assessment methods.

The censored approach may also be relevant when management regulations affect the interpretation of fishery data. For example, changes to bag limits can affect CPUE. It seems possible to treat the bag limits as censoring variables, and estimate CPUE including the impacts of bag limits.

It was suggested in the presentation that unsubstantiated uncertainty should be challenged, and I agree. The flip side is that unsubstantiated certainty should also be challenged. Informative Bayesian priors can fall in this category. Punt and Hilborn (1997) noted a tendency, although they were not clear by whom, to underestimate uncertainty, and hence to specify unrealistically informative priors. Informative priors need to be rigorously challenged. As a rule, priors should not be based on the assessment model data because this is essentially 'using the data twice' which leads to a false measure of precision that is too high. An extreme example of this is using the posterior for parameters from the last assessment as priors in the current assessment. Clearly, after repeated application of this strategy the assessment would indicate very precise information about parameter values.

<u>Bayesian uncertainty techniques as applied to sharks</u>

Much of the talk dealt with specific issues, and not enough detail was presented for me to review the specifics. It seemed to be a work in progress.

However, the issue of model complexity was raised. How complex do we need to be? This is relevant for estimating uncertainty. While over-simplified models may give false and too precise measures of uncertainty, overly complicated models may do the same. It is well known that maximum likelihood estimates of variance parameters are biased low, and the bias can be substantial in highly parameterized models. This leads to confidence intervals for important parameters that are too narrow. This problem can also affect bootstrap methods. The extreme case is the very highly parameterized model that fits the data exactly, and produces zero residuals. Clearly bootstrapping in this case produces false precision, and something similar happens in less extreme, but still highly parameterized, models. Various adjustments have been proposed to deal with this bias problem (e.g. Severini, 2000) but as far as I am aware this problem has not been investigated for stock assessment models. However, this is an important reason why assessment scientists should verify that their confidence intervals are not biased even if their modeling assumptions are true. A similar problem likely affects Bayesian posteriors when they are interpreted in the frequentist perspective.

Gulf of Mexico assessment uncertainty methods

Uncertainty in assessment model formulation and model inputs was often addressed using sensitivity analyses. The presentation gave a good overview of the types of model formulation uncertainties that occur. For reasons outlined above (see my comments for South Atlantic assessment uncertainty methods) I do not think that ad hoc sensitivity analyses give a measure of uncertainty that is credible or relevant. A better approach is to incorporate model uncertainty using process error, or to utilize more realistic error assumptions about model inputs.

If sensitivity analyses are provided then the assessment workshop (AW) should also provide some quantification of the likelihood or plausibility (i.e. weightings) of each scenario. AW participants will often be the most informed group to do this. If the AW does not provide the scenario-weightings then others, who may be less informed about the stock, will have to do it.

For example, the assessment for Gray Triggerfish in SEDAR9 used three scenarios for M: 0.25, 0.27, and 0.3. This seems to be a narrow range; is M=0.23 not plausible, less plausible, equally plausible? These sensitivity analyses to measure the M axis of uncertainty may under-estimate uncertainty.

In the case where the assessment information provides little information about values for M, I suggest a better approach to deal with uncertainty about M is to simply treat M as random with some distribution that reasonably reflects uncertainty about its value. The distribution would have a fixed mean (e.g. 0.27) and CV that gives reasonable weighting to other values for M. This would replace the need for sensitivity analyses for M, and provide an evaluation of stock status relative to benchmarks that incorporates uncertainty about M. Note that this approach is different than the Bayesian approach in which a prior distribution is updated based on the data. In the approach I suggest there is no updating because the assessment data provides little information about values for M. I think using a Bayesian approach in this situation can produce spurious results.

A similar approach could be used to address uncertainty about steepness or landings. However, if the assessment data are informative about some of these factors then this information should be used so that statistical inferences are relevant. Sensitivity analyses could be used to first determine if the data provide information about a parameter. If they do, then a Bayesian or Empirical-Bayesian approach could be used; but if the data are uninformative about the parameter, then the subjective “sensitivity distribution” approach could be used. The sensitivity runs are used to defend the assessment model formulation but are not used to characterize uncertainty.

In the presentation it was mentioned that including only recruitment uncertainty in projections is insufficient, especially for short-term forecasts of fisheries that are not heavily dependent on recruitment (i.e. not a recruitment fishery). In this case it is clearly important to include uncertainty about the start position in projections (i.e. estimation error in survivors). It is also important to include process error and catch estimation error although it may be difficult to separate these sources of variability.

It is becoming more standard to incorporate process error in population dynamic models to reflect natural variability not accounted for by the model, and thus to increase the relevance of the model. It is necessary to account for process error for realistic (i.e. appropriate) stock projections. Process error should be included in the calculation of equilibrium reference points and can have important effects. Harvesting according to the deterministic MSY rule is an underoptimized strategy and can lead to strong decreases of the resource (e.g. Bousquet et al., 2008) when there is process error. The deterministic MSY exploitation rate is incompatible with the assumption of equilibrium: on average, one cannot hope to harvest more than the stochastic MSY. Constant harvesting at the deterministic MSY exploitation rate will eventually lead to stock extinction.

Beaufort approach

There is a problem with invoking asymptotic arguments to justify statistical inferences with most stock assessment models. In the regression setting for generalized linear models, the basic condition underlying asymptotic normality is that the "information" about all parameters grows as the sample size increases (e.g. Fahrmeir and Kaufmann, 1985). Note that simply increasing sample size does not ensure that the information increases. Something similar must apply to fisheries models.

In a standard ADAPT-type VPA the information about survivor parameters does not grow as sample size increases (i.e. as the number of years increase). Estimates of survivors are derived basically from the last few survey observations and estimates of survey catchability (Q). Even if a very long time-series is available, the estimates of survivors will still be uncertain, although estimates of Q will be highly certain. In general, asymptotic arguments may not apply to parameters with year subscripts. This applies to the asymptotic normal approximation of the distribution of the mle, to the chi-square approximation to the profile likelihood, and possibly to the validity of bootstrap confidence distributions. Unfortunately it seems that statistical inferences for current stock size will not enjoy the distributional robustness afforded by asymptotic arguments. This is an area that requires additional research.

Bootstrapping recruitment residuals for quantifying benchmark uncertainty was discussed. A difficult issue here is auto-correlation in recruitment residuals that is often apparent.

Additional process error in stock dynamics, and implementation error in regulating catches, is also important to consider when deriving benchmarks. If actual catches vary from the target or if there is process error in the model then deterministic MSY catches will not be sustainable in the long term.

The Inverse Prediction approach that was described seemed complicated, especially when parameter co-linearity is present.

ToR 2: Review and discuss alternative approaches to characterizing and presenting uncertainty in stock assessments and projection analyses, including those utilized in other regions. Discuss the sources of uncertainty which require consideration in the stock assessment process. Sources may include implementation uncertainty, within model uncertainty, inter-model uncertainties. Make recommendations on which sources of uncertainty to consider for future SEDAR stock assessments and projection analyses.

The following presentations related to ToR 2 were reviewed. I provide the abstracts to help motivate my comments, which are provided at the end of this section.

<u>PIFSC assessment uncertainty methods. J. Brodziak, PIFSC, Honolulu.</u>

The Pacific Islands Fisheries Science Center (PIFSC) conducts stock assessments of insular and pelagic fisheries resources in the North Pacific Ocean. In this presentation, the primary methods used by the PIFSC to characterize uncertainties in stock assessments are described using specific examples. Major sources of uncertainty for the stock assessment process include: (1) model uncertainty/structural complexity, (2) estimation error, (3) sampling/observation error, (4) natural variability/process error, (5) implementation uncertainty, and (6) inadequate communication among scientists, managers, and stakeholders. The Western and Central Pacific bigeye tuna (*Thunnus obesus*) stock assessment was presented as an example of incorporating model uncertainty, estimation error, and communication uncertainty into the fishery system analysis. This bigeye tuna stock is a data-rich stock that is heavily exploited. This point was emphasized through extensive analyses of alternative reference cases (i.e., plausible assessment model scenarios) and graphical presentation of comparative results to the Western and Central Pacific Fisheries Commission, the regional fishery management organization. The Hawaiian bottomfish stock complex, an insular management unit comprised of deep-water snapper and grouper species, was presented as an example of a data-poor stock assessment in the Pacific where parameter uncertainty was estimated using Markov Chain Monte Carlo (MCMC) simulation in a fully-Bayesian stock assessment model. In addition, the implementation uncertainty about the effectiveness of new restricted fishing areas (i.e., MPAs) on this stock complex was addressed using a conceptual model of fishermen's behavior in relation to the redistribution of fishing effort inside MPAs and the potential for noncompliance with MPAs. This analysis suggested that redistribution of effort and noncompliance could compromise the effectiveness of the new MPAs and this information was considered by the Western Pacific Regional Fishery Management Council in selecting seasonal closures to reduce bottomfish fishing mortality. The North Pacific swordfish (*Xiphias gladius*) stock assessment was presented as an example of how structural uncertainty was incorporated into the stock assessment and management advice using two alternative stock structure assumptions, the two-stock and the single North Pacific stock scenarios. This data-moderate assessment also provided an example of how implementation uncertainty was included in management advice through uncertainty in projections of future fishing mortalities and associated catches using MCMC results that were presented to the International Scientific Committee for Tunas and Tuna-like Species in the North Pacific. The North Pacific striped marlin (*Tetrapturus audax*) stock assessment provided another example of a data-moderate assessment where structural uncertainty was incorporated into management advice through the inclusion of two equally-plausible alternative stock-recruitment steepness scenarios. This data-moderate stock assessment also included evaluation of parameter uncertainty about biological reference points and steepness uncertainty through the application of model-averaging to calculate biological reference points. Overall, the characterization of stock assessment uncertainty in the North Pacific is currently accomplished by a variety of methods and is an active area of research for both insular stocks and highly-migratory pelagic species.

<u>NEFSC assessment uncertainty methods. C. Legault, NEFSC, Woods Hole.</u>

The NEFSC assesses 50 species comprising over 60 stocks using a number of standard assessment models from the NOAA Fisheries Toolbox mainly. The choice of model usually depends on the available data, with data limited stocks using AIM, CSA, or SCALE while data rich stocks use VPA or ASAP, generally. Uncertainty in the assessment models is estimated either through bootstrapping residuals of index fits (AIM, CSA, VPA) or through Markov Chain Monte Carlo methods (SCALE, ASAP). The model results are often used as uncertain starting points for projections using AgePro, which also incorporates uncertainty in future recruitment. Yield per recruit is often used to set the F reference point while AgePro is used to project this F many generations into the future to estimate MSY and Bmsy, such that the reference points and projections will be consistent.

Time series plots and probability density functions are commonly employed to graphically display the level of uncertainty in assessments. In some cases, risk plots are generated which show the probability of exceeding an F reference point or of biomass not increasing by a given amount for different levels of projected catch. Some stocks have multiple models recommended for use with management decisions when a single model cannot be agreed upon. These multiple models demonstrate the among model uncertainty, but can lead to difficulties in setting quotas.

Retrospective patterns have become a standard diagnostic for NEFSC stock assessments and cause an additional level of uncertainty when present. Not all stock assessments in the Northeast exhibit a retrospective pattern, and not only one model type will exhibit a retrospective pattern. Splitting the survey time series has been identified as one possible “fix” to retrospective patterns and preliminary MSE results demonstrate that it is effective even when misreported catch or changes in M are the real cause of the retrospective pattern. Comparison of alternative retrospective fixes, such as modifying the results of the base model by the estimated amount of retrospective patterning, often result in similar stock status and catch advice.

A new uncertainty has been introduced to stock assessments in the Northeast by the deployment of a new research vessel using new trawling gear. Calibration coefficients are being estimated to allow linkage of the data collected with this new system to the 40+ year time series. Currently beta-binomial conversion coefficients have been estimated for catch/tow in numbers and weight. Exploration of length-based conversion coefficients is continuing. One advantage of this exercise has been the ability to set a prior on the catchability coefficient for the old vessel and gear based on the conversion coefficient. This prior is used with minimum swept area abundance estimates from the survey to help determine the magnitude of population abundance. This example demonstrates that there will always be new sources of uncertainty in stock assessments over time and the analysts must be prepared to address them as they are encountered.

<u>SWFSC assessment uncertainty methods. R. Conser, SWFSC, La Jolla.</u>

The Southwest Fisheries Science Center (SWFSC) carries out stock assessment research in four general areas:

1) Species managed under Fishery Management Plans (FMP) developed by the Pacific Fishery Management Council (PFMC);
2) Pacific highly migratory species (HMS) under the purview of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific (ISC);
3) Protected Species (marine mammals and sea turtles); and
4) Antarctic Species (krill, crab, and finfish including effects on seals and seabirds).

This paper focuses on the first of these areas as it most relevant to the terms of reference for the SEDAR Procedural Workshop on Uncertainty.

The PFMC has four FMPs: Groundfish, Salmon, HMS, and Coastal Pelagic Species (CPS). Stock assessment scientific support for the PFMC is provided jointly by the SWFSC and the Northwest Fisheries Science Center (NWFSC). Scientists from both Centers conduct stock assessments and serve on the PFMC SSC and its FMP-specific management teams. A major research effort over the past year has focused on analyses needed to support the amendment of all FMPs so that they conform to the new National Standard 1 (NS1) guidelines. This work – led by the SSC – has mainly dealt with species in the Groundfish and CPS FMPs to date.

The NS1 guidelines require specification of the annual overfishing level (OFL) catch and an allowable biological catch (ABC) for each managed species. ABC must be less than OFL – the difference being a function of the scientific uncertainty and the probability of overfishing. Determination of the OFL buffer (ABC/OFL) is a joint SSC and Council responsibility. The SSC quantifies the scientific uncertainty (σ) and the Council decides the acceptable probability of exceeded OFL (p^*). The latter must be less than or equal to 0.50 but otherwise is driven by the Council's level of risk aversion, cost factors, and other management related issues. In all cases, OFL buffers should be larger for data-poor stocks than for data-rich stocks.

The SSC has conducted analyses aimed at i) estimating the level of σ for a wide variety species in the Groundfish and CPS FMPs and ii) developing an algorithm that determines the OFL buffer as a function of σ and p^*. The details are summarized in a companion paper in this volume (An Approach to Quantifying Scientific Uncertainty in West Coast Stock Assessments).

Preliminary results show that for 16 groundfish and coastal pelagic species, the mean σ on terminal biomass is 0.19. This represents the average amount of statistical measurement error within assessments conducted for the PFMC. In contrast, the average σ ascribable to model specification error (i.e. among assessment variation) is 0.34, which is the far greater of the two sources of uncertainty. An example from the preliminary results follows: if only among assessment variation is considered; if the variance in FMSY is ignored; and if p^* is fixed at 0.40, the OFL buffer would be 0.92, i.e. an 8% reduction in harvest from the OFL level. Smaller p^* and/or introduction of additional variance components (e.g. within assessment or FMSY variance) would reduce the harvest further.

The PFMC SSC's developmental work on this method is continuing.

<u>AFSC assessment uncertainty methods. D. H. Hanselman, AFSC, Juneau.</u>

Alaska fisheries management is a fundamentally conservative framework that sets quotas based on a set of tiers of data availability. The amount of data available is directly related to the amount of uncertainty in those stocks. The framework also relies on the "lowest common denominator," meaning that any Overfishing Limit (OFL) that is exceeded can constrain the rest of the fisheries that catch that stock. The North Pacific Management Council (NPFMC) administers this framework for groundfish on an annual basis under the axiom that fisheries are only allowed to develop only when sufficient data has been collected. Uncertainty enters the process at several levels. In the tiers with statistical age structured models (1-3), uncertainty is accounted for in the modeling process through the error assigned such things as survey biomass estimates, ageing error, catch error, and prior distributions assigned to key parameters. In Tier 1, acceptable biological catch (ABC) is assigned explicitly based on uncertainty. This is done by using the harmonic mean of the fishing mortality rate and the geometric mean of exploitable biomass. These means are always lower than the arithmetic mean, and how much lower is determined by the variance of model outputs. Only a few of the most data rich stocks are in this category where a stock-recruitment relationship and MSY are estimated. Most of the target stocks in Alaska are administered under Tier 3, where proxies for MSY and OFL are determined by spawners-per-recruit reference (SPR) reference points. In this tier, uncertainty in those reference points is compensated for by setting maximum ABC at an SPR rate below the MSY proxy rate. This usually yields a buffer from 15-20% below the MSY proxy. For Tiers 1-3, a control rule is in place that reduces fishing mortality when stocks fall below reference points.

In general non-target stocks are located in Tiers 5-6. These species are caught in fisheries, but are considered non-target or incidental catch species. In Tier 5, OFL is set at F=M and in Tier 6, OFL is set to average catch for a specified period. The reference points buffer against uncertainty by setting ABC at 0.75 these levels. Generally, a fishery is not allowed to develop on these stocks until sufficient data is collected to increase their Tier level. Retrospective analyses are conducted to further assess model uncertainty. Projections are done stochastically including recruitment uncertainty, as well as full posterior estimates of projection uncertainty. Ecosystem considerations are often taken into account as a source of uncertainty, and are commonly used to provide additional inputs to the quota setting process as support for precautionary reductions.

For groundfish management in Alaska, the current buffers and control rules may be sufficient to meet the spirit of National Standard I guidelines, but further analysis may lead to changes to increase the level and rigor of scientific uncertainty included in these buffers. One way to do this which would apply to nearly all Alaska stocks is to relate the uncertainty of OFL to survey uncertainty for use in the P* method. For crab and scallop management, no ABC control rule is currently in place and the P* method is being proposed to create this rule. Different levels of uncertainty for OFL are proposed by using strictly model uncertainty, or an additional level added in from retrospective analysis or more ad hoc approaches of adding a constant like 0.2 or 0.4 to the coefficient of variation. Future work will include management strategy evaluations to determine the robustness of current and proposed uncertainty methods.

ICCAT methods. M. Ortiz, SEFC

Since 1965, the ICCAT regional fisheries management organization has assessed the main tuna and billfish species of the Atlantic Ocean and Mediterranean Sea including: bluefin, bigeye, yellowfin, albacore, swordfish, skipjack, blue and white marlin, and sailfish. Recently, the ICCAT Standing Committee on Research and Statistics (SRCS) has initiated assessments of other pelagic species including sharks and bycatch species that interact with the main gears of tuna operations, such as seabirds.

Main sources of uncertainty in the assessment of these species can be classified into:

Catch data
- Directed Landings
- Non-targeted/Bycatch (Uncertainty generally greater than for the directed landings)

Biological and population information
- Limited biological sampling/studies with low coverage in space and time
- Low sampling for size/age composition
- Very limited scientific or non-fishery surveys

Assessment Models
- Models used by the SCRS include, Age structured analyses (restricted to main tuna species) and surplus production models (data-poor species). Recently, the SCRS has introduced Statistical Catch at age models (Multifan-CL, Stock Synthesis) but still they are not the main models for management advice
- Structural uncertainty is usually assessed using sensitivity analyses of alternative model formulations
- Within-model uncertainty generally is assessed using bootstraps, particularly from tuning indices. Annual trajectories of stock status are generally illustrated using phase plots base on the bootstrapped results.

Stock projections
- Stock recruitment assumptions for short term projections and definition of stock benchmarks
- What Sources of uncertainty of the final model should be carry over

Management implementation
- ICCAT convention objective is MSY but there is not a formal distinction between a target objective and limit threshold(s).

The SCRS has reviewed these sources of uncertainty, and make recommendations to the Commission that resulted in specific task to reduce uncertainty. For example, improvements in data submission and compliance, by providing economical and technical assistance to developing countries to improve data collection and sampling, initiation of species-specific programs for biological studies, improved observer sampling of major fisheries in the Atlantic, the creation of a Precautionary Approach Working Group, and the adoption and implementation of methodologies used by other RFMOs to communicate uncertainty in assessment results. The Commission is also revising its objectives in order to adopt a more risk adverse policy of fisheries management. The presentation provided an example of a preliminary evaluation of potential limit benchmarks applied to the northern albacore stock and the implications of catch under-reporting uncertainty in the projections of eastern bluefin tuna.

Ralston/PFMC approach, C. E. Porch

Quantifying scientific uncertainty in estimating an appropriate catch level for a fish stock is challenging. Multiple sources of error can easily be identified, including measurement error that is conditioned on the adopted model, model specification error, forecast error, and uncertainty about overall stock productivity. In addition, there are without doubt other unknown factors that will negatively influence the precision of scientific advice on catch levels. Notwithstanding these difficulties, the Magnuson-Stevens Reauthorization Act (MSRA) identifies the quantification of scientific uncertainty in the development of advice on catch levels as a key requirement of the new law. Moreover, the Scientific and Statistical Committees (SSCs) of the Regional Fishery Management Councils have been given the responsibility to quantify that uncertainty.

While many sources of uncertainty exist, the focus here is on quantification of statistical measurement error and model specification error, particularly the latter. While not all inclusive, the study of these two factors is feasible with the information that is currently available. They are also likely to include the dominant sources of scientific uncertainty in the development of scientific advice vis-a-vis groundfish and coastal pelagic species catch levels at the Pacific Fishery Management Council.

Although full Bayesian integration through MCMC calculations is a preferred method of estimating measurement error “within” a stock assessment, an inadequate number of studies have successfully achieved that type of analysis. Consequently, we report the first order approximate estimates of the standard error on terminal biomass from stock assessments that are calculated by inversion of the model’s Hessian matrix (i.e., the asymptotic standard error). To summarize variation “among” stock assessments, as a proxy for model specification error, we characterize retrospective variation among multiple assessments of the same stock.

Preliminary results show that for 16 groundfish and coastal pelagic species the mean of the coefficient of variation on terminal biomass is 0.19. This represents the average amount of statistical measurement error within assessments conducted for the PFMC. In contrast, the average coefficient of variation ascribable to model specification error (i.e., among assessment variation) is 0.34, which is the far greater of the two sources of uncertainty. Given the preliminary results: if only among assessment variation is considered; if the variance in FMSY is ignored; and if the probability of overfishing is fixed at 0.40, an appropriate OFL buffer on the overfishing catch level is to reduce the harvest by approximately 8%. Smaller acceptable probabilities of overfishing would reduce the harvest further.

The PFMC SSC's developmental work on this method is continuing.

Reviewer's views

The review part of this ToR was fairly well addressed. Most of the other regions in the US gave presentations on how they accounted for and incorporated uncertainty in their assessments. The review was weak in terms of what is going on outside of the US (i.e. only an ICCAT presentation), but I think little innovation was missed in terms of Canadian or ICES approaches. I conclude that other regions did not demonstrate that they are accounting for model or implementation uncertainty much better than in SEDAR. An empirical approach to account for model specification uncertainty was proposed, but I conclude that further research is required before one could consider it to be a substantial improvement (see comments below for Ralston/PFMC approach).

A succinct summary of important uncertainties in stock assessment was not reviewed at the workshop. This may yet appear if the workshop report. I provide a partial summary in **Appendix 4**.

The workshop report will describe many possible sources of uncertainty. The relevance of these sources of uncertainty will vary from stock to stock; however, some sources of uncertainty like M and catches will be relevant for many stocks. The workshop did not specify a "must do" list of sources of uncertainty to consider for future SEDAR stock assessments and projection analyses.

The following comments were motivated by the presentations and so I present them in that format; however, some of the comments are fairly general. Many of my comments (e.g. sensitivity analyses, asymptotics, stochastic MSY) related to presentations made for ToR 1 also apply here, and are not repeated.

PIFSC assessment uncertainty methods

Profile likelihood confidence intervals for high parameter models were advocated. They have some improved properties. They can be asymmetric and do not cover infeasible (e.g. negative) values. However, profile likelihood confidence intervals only have the same order of accuracy as standard linear approximation normal methods. The Chi-square repeated sampling approximation of the profile likelihood statistic may not be good. The distribution of this statistic is known to be F or non-central chi-square in some situations.

NEFSC assessment uncertainty methods

Retrospective patterns usually occur when there are time trends in model residuals, but these patterns can also occur when there is confounding between index catchability and stock size. In the first case, retrospective patterns indicate a problem with some model assumption. However, the lack of retrospective patterns does not mean there are no problems with model assumptions.

If there is uncertainty in reference points then short term risk plots should be relative; that is, plot Pr(Bproject/Bref > 1|TAC) and Pr(Fproject/Fref > 1|TAC).

A change in research vessels has been a big issue for recent NEFSC assessments. It would be desirable to account for the uncertainty in vessel/gear calibrations in stock assessments. Hopefully this means that more recent estimates of trawlable abundance are more precise than the older estimates. However, this may not translate into decreased variability in stock size indices. The stock assessment model index variability involves measurement error in trawlable abundance plus additional variability related to the fraction of the stock available to the survey (i.e. coverage error). The survey provides no information about coverage error which may change over time as abundance changes. Hence, total error could be greater since a gear change, even if the measurement error component has decreased.

SWFSC assessment uncertainty methods

The method of choice for incorporating uncertainty into management advice is the delta method. This has deficiencies outlined under ToR 1.

It was demonstrated that among assessment uncertainty has been larger (retrospectively) than within assessment uncertainty. This partially indicates that the methods used in the past to measure within assessment uncertainty did not work.

AFSC assessment uncertainty methods

The presentation suggested that it is important to include within-survey variance in stock assessment models, and I agree. However, it is not straight-forward how to do this. The within-survey variance is a component of the index variance in the assessment model. Another important component of variability that the within-survey variance does not include is related to the fraction of the stock available to the survey (i.e. coverage error). The survey provides no information about coverage error.

Ralston/PFMC approach

The approach to quantifying stock assessment error assumed that the model specification error was constant over time and could be estimated using historic assessment results. There are many reasons why this assumption may not hold in particular cases. For example, early assessments may have been based on short time series in which stock size and index catchability were confounded. This tends to make the assessments more sensitive to model mis-specification. More recent assessments may not suffer from this problem as much, and model error could be less. Also, it may be that important new data are available (e.g. tagging) that provide information on parameters (like M, selectivity) that was not available for earlier assessments. This would also reduce model error. Additional comments are presented in the Group 3 section of ToR 3.

A way to empirically test the approach advocated to determine model specification error would be to give the current assessment data (not stock name, but basic biological information) to many assessment experts and examine the range of assessment advice produced.

ToR 3: Recommend approaches for representing uncertainty in the assessment documentation and expressing confidence in estimated parameters. Discuss both inter- and intra-model uncertainty. Include guidance for different model classes.

Reviewer's views

There was some uncertainty at the workshop about how to respond to the ToR's, which were somewhat vague. It could be argued that the ToR's were too broad to be competently addressed in a one week meeting. The ToR's were more suitable for a multi-year working group. Also, the theory and technology for representing uncertainty is still evolving which makes it difficult to recommend best practices, especially when the efficacies of the various approaches are not well understood.

The workshop did not make specific recommendations for expressing confidence in estimated parameters. Instead, a range of options were discussed. There was good discussion on inter- and intra-model uncertainty, but only a range of options was presented for accounting for this. It seems that assessment teams must still determine for themselves the best way to achieve this.

The workshop steering committee decided to split the participants into four subgroups dealing with: 1) Data. Construct a checklist of data types and the types of uncertainty that need to be included in assessments. 2) Models. Construct a checklist of assumptions, model types, and sensitivities. 3) Output. How to derive a distribution for OFL? 4) Implementation uncertainty.

The following are my comments on the progress reported by the four subgroups.

<u>Group 1 (Data)</u>

The workshop developed a spreadsheet to characterize and qualitatively rank data sources with respect to uncertainty. Comment boxes were also available to indicate special elements in the data not readily represented by the ranking score and to indicate reasons for rankings Temporal and spatial coverage or limitations could also be indicated in the comments field.

This could be a valuable tool for representing and documenting uncertainty in the assessment, but it will require some iteration. I was concerned if the rankings would be comparable across stocks, and if users will interpret the rankings as intended. This will be discovered through trial and error, so it seems sensible to continue development of the spreadsheet using a small number of case studies. An important goal should be to provide as much objective information as possible to assist in specifying prior distributions. The people who will fill the spreadsheet will, in many cases, be the most knowledgeable for specifying prior distributions.

However, it is important to keep in mind that most assessment model estimation methods that use priors will assume that the priors are independent of the assessment data. Technically this means that the priors should not be specified using the assessment data. Otherwise, model estimates will be too precise. In practice this will be a difficult issue. If the data are used weakly in some respect then this may not be an important issue. For example, if longevity information is derived from survey catches to assist in specifying M then this may not be an important dependency. Alternatively, if values for carrying capacity are based on average catch then this would be an important dependency. Empirical Bayes models estimate priors from the data, and that is OK because this is taken into account in statistical inference.

<u>Group 2 (Models)</u>

Guidance was provided on the choice of analytical tool or model, which depends on the types and quality of data available. The group advocated that the appropriate model is one that is capable of integrating all of the available relevant and informative stock data. The group provided a review of the most important sources of uncertainty in stock assessment models. They divided uncertainty into two major sources: parameters and model structure. A more complete review of parameter uncertainty is given in **Appendix 4**. Structural uncertainty is more amorphous. Examples given included stock structure, the start year of the model, migration patterns, and the spatio-temporal definition of the data. Model averaging techniques were recommended to incorporate structural uncertainties. It was recommended to account for parameter uncertainty using stochastic variables (e.g. M) and/or priors in a Bayesian approach.

The guidance was very high level and will not have much traction with assessment groups unless it can be made more practical. I did not get the sense from assessment scientists at the meeting that they had a clear understanding of how to implement recommendations. Case studies would have been a useful way to demonstrate how to implement recommendations, although clearly time at the workshop was a limiting factor.

One can also expect much “devil in the detail” in terms of implementing the recommendations. For example, the bootstrap was recommended as an approach to consider for characterizing within model uncertainty. However, no guidance was given about how to do the bootstraps (e.g. parametric, nonparametric) for indices or size compositions, and how to form confidence distributions (e.g. bias corrected, accelerated bias corrected, bootstrap-t, etc). The delta-method using the Hessian matrix or likelihood profiles was mentioned, but no guidance was given about what to do in models with many parameters when asymptotic distribution results are suspect. This level of advice was beyond the capacity of the workshop, but within the scope of the ToR.

This group did not provide any guidance on how approaches to account for uncertainty may differ from different model classes.

<u>Group 3 (Outputs)</u>

The group identified two basic approaches to characterizing uncertainty in the outputs of stock assessment models; (1) an empirical approach that used existing historic information on the consistency of stock assessment outputs through time, perhaps averaged across groups of stocks and (2) a model based approach to account for sources of uncertainty in the envelope of total uncertainty.

I provided some comments on the empirical approach (see <u>Ralston/PFMC approach</u> in ToR2). This approach is unconcerned with the validity of assessment assumptions (primarily reflected in fits to the data). It was not clear on how the approach would apply when multiple models (e.g. decision table) are used for advice. The empirical approach may not explicitly accommodate known sources of uncertainty (e.g. M, catches, etc), only unless such sources of error manifest themselves as variation in the assessment of stock status over time. For example, if all historic assessments assumed M was 0.2, and there is uncertainty about this value, then how is this reflected in the empirically derived variance? If none of the historic assessments included observation error in input data then how does the empirical variance account for this? It seems feasible that the empirical variance is partially reflecting parameter estimation error, process error, and a component of model specification error.

I feel the empirical approach requires further study. At present it seems to be an ad hoc method whose efficacy is a matter of faith. I have no idea how the method could be validated theoretically or more generally in a scientific manner. Perhaps the first step is to see the approach published in a scientific journal. If the approach could be simulation tested in some meaningful way then this would also help. However, at present I could not defend the empirical approach amongst my scientific peers so I do not recommend its usage.

Model based approaches were also considered. The discussion of model based approaches overlapped considerably with Group 2 discussions, although Group 3 considered model averaging techniques in a little model detail. Specific recommendations were not provided. The group (i.e. workshop) suggested that the primary advantage of model based approaches is that they can incorporate the most up-to-date information and assessment methods. A disadvantage is that it is difficult to discern whether the model adequately accounts for the major sources of uncertainty including uncertainty about model structure, parameter estimation error, observation error, and inherent natural variability (process error). The group suggested that some important sources of uncertainty could be accounted for using well-thought out sensitivity runs, i.e., a set of "alternative states of nature". One could simply present the results of alternative states of nature as ancillary information, perhaps via a decision table, or one could apply model-averaging techniques, where "plausibility" weights are assigned to each candidate model, perhaps based on some measure of the fit to the data, e.g., AIC, inverse-variance weighting, or expert opinion.

The spreadsheet developed by group 1, when finalized, will make it much easier to discern if the model has adequately accounted for observation error in inputs. It was not clear what was meant by "parameter estimation error". If this also referred to assessment model inputs (like M) then the spreadsheet will help here as well.

Model uncertainty is a difficult issue. It will often be difficult to select a balanced set of sensitivity runs to reflect uncertainty. For example, if there is error in catch, do runs with twice the observed catch or half the observed catch give balanced alternative states of nature. The answer to this question is linked to whether a 50% reduction in catch is the same size error as a 100% increase in catch. Many assessment scientists will simply not know the answer to this. Hence, if there are three alternative states of nature, based on assessments with catches equal to 0.5×, 1.0×, and 2.0× reported values, then what plausibility weights should be assigned to the scenarios? I think it will be hard to select from the weights (30,40,30), (25,50,25), (10,20,70), etc. Hence, there will be substantial uncertainty in subjective weights used for model averaging. Also, errors in catches will vary by age and year, and the chosen alternative states of nature may not adequately reflect this uncertainty. It may be that the largest uncertainty in inferences about stock status comes from trends in catch errors rather than simple scalings. This can be investigated using more sophisticated perturbation analyses (e.g. Cadigan and Farrell, 2002); however, the basic problem still remains about how much weight to give to different perturbations.

Using model fits to assessment data to determine weights may also have problems, including robustness. Models may fit better for spurious reasons. For example, in a stock assessment I was responsible for, estimated survey catchabilities (Q's) had a domed pattern and when Q's were made flat then the assessment model did not fit the data as well. However, the improvement in fit by not assuming flat Q's came from only three ages in one year of a survey index which was strongly suspected to be a year effect because of an extremely large catch that year. Weighting these two scenarios by models fits does not seem appropriate.

It is probably not useful to give specific recommendations on how to account for model uncertainty. However, if the sources of model error are random variations in input parameters (i.e. M = 0.2 + error) and data (i.e. catches = reported + error) then I do think there is good potential that models with process error can, in aggregate, account for these uncertainties. If the mean of the errors is not zero then adding process error will probably not be sufficient.

Group 4 (Implementation uncertainty)

This group developed an amazingly long list of factors that effect implementation uncertainty. They highlighted difficulties (perhaps futility) in managing fisheries. They did not provide recommendations on how to account for or incorporate these uncertainties in stock projections.

ToR 4: Review and discuss uncertainty estimates needed for management, specifically addressing the needs for each council's ABC control rules and ACT determinations. Make recommendations on which uncertainty estimates should be included in future SEDAR stock assessments and projection analyses.

The following presentations related to ToR 4 were reviewed. I provide the abstracts to help motivate my comments, which are provided at the end of this section.

Gulf of Mexico SSC control rule presentation, R. S. Fulford, GMFMC Standing SSC

The Gulf of Mexico SSC formed an ABC Control Rule Working Group in July of 2009 that is comprised of members of the standing SSC, members of the GMFMC, and council staff. After review of several methods currently in use by other regional councils for setting ABC, the working group has adopted the p* approach as the primary method for determining the size of the buffer between OFL and ABC for each individual assessment. The working group is currently working to develop scoring criteria for four dimensions that will result in a range of p* between 0.05 and 0.35 that when applied to the pdf for OFL will result in a range in p(overfishing) between 0.45 and 0.15. The 0.15-0.45 range is based on guidance regarding acceptable risk provided by the GMFMC. The four dimensions currently under review by the working group are 'Data quality', 'Characterization of data uncertainty', Characterization of process uncertainty', and 'PSA analysis.' The exact form of the p* calculation for GOM stocks has not been finalized. In addition the working group is also working on methods for data poor species, which they characterize as any stock for which both an direct estimate AND a pdf for OFL are not available. In these data poor situations the working group has proposed a decision tree approach based on a method currently used by the PFMC for calculating ABC for groundfish stocks. This approach involves a sliding scale from the purely quantitative approach (i.e. p*) to a purely policy based approach (e.g. x * average catch) that allows for the use of the maximum amount of data available for any individual stock. This data poor decision tree has only been developed in preliminary form, but we anticipate a completed draft will be ready for GMFMC review by late spring 2010.

South Atlantic SSC control rule presentation. C. Belcher, SAMFC Scientific and Statistical Committee Chair

The South Atlantic Fishery Management Council's Scientific and Statistical Committee convened a special meeting in March 2009 to focus on the development of an Acceptable Biological Catch (ABC) Control Rule for South Atlantic fish stocks. The control rule assumes that an estimate of the overfishing level (OFL), stated in weight, has been calculated and some reasonable measure of statistical uncertainty about the OFL also exists. The concept developed by the SSC is designed to be objective in nature and operates by adjusting the probability of overfishing or P* value. The control rule generates penalties or reductions based on four characteristics of stock assessments: Assessment information, Characterization of Uncertainty, Stock Status, and Productivity and Susceptibility. The assessment information dimension reflects available data and assessment outputs. The characterization of uncertainty dimension reflects how well uncertainty is characterized in the assessment, not the actual magnitude of the uncertainty. Stock status is included among the dimensions so that an additional adjustment to ABC can be added for stocks that are overfished or overfishing. The final dimension addresses biological characteristics of the stock, including productivity, which reflects a population's reproductive potential, and susceptibility to overfishing, which reflects a stocks propensity to be harvested by various fishing gears. Each dimension has a maximum penalty of 10% associated with it. The sum of penalties is subtracted from the base case of P*= 50%, which is when the ABC=OFL. Depending on the characters and results of a given stock assessment, the corresponding P*, which is used to determine the ABC value, can range from 50% to 10%.

Reviewer's views

Both presentations reviewed ACT determinations and outlined the needs for ABC control rules. The two controls rules developed for "data rich" stocks were similar, and generate penalties or reductions based on four characteristics of stock assessments: Assessment information, Characterization of Uncertainty, Stock Status, and Productivity and Susceptibility. However, each control rule utilized different scoring mechanisms. I did not have any sense of the strengths or weakness of these differences.

Each control rule adjusts the probability of overfishing based on an assessment of the quality of the stock assessment. I did not understand the motivation for the approach. Why should the probability of overfishing vary from stock to stock? Some of the dimensions should be reflected in the OFL CV. However, one would not want to double-penalize for uncertainty.

Nonetheless, each presentation clearly outlined the information that will be required to evaluate the control rules. The Gulf of Mexico ABC Control Rule was identified as a work in progress, and hence its needs are as yet not definitive.

The Gulf of Mexico ABC Control Rule presentation included consideration of data poor species, while the South Atlantic SSC control rule did not.

I have little experience in evaluating harvest control rules; however, it may be useful to examine the potential impact of these control rules using MSE.

The workshop did not make specific recommendations on which uncertainty estimates should be included in future SEDAR stock assessments and projection analyses.

ToR 5: Provide recommendations on best use of uncertainty characterization and recommendation methods such as P analysis, risk evaluation approaches, PSA.*

The following presentations related to ToR 5 were reviewed. I provide the abstracts to help motivate my comments, which are provided at the end of this section.

p* approach. K. W. Shertzer, SEFC, Beaufort.

In U.S. federal fishery management, acceptable biological catch (ABC) is set below (or equal to) the overfishing limit (OFL) to account for scientific uncertainty, and annual catch targets (ACTs) are set below (or equal to) the ABC to account for implementation uncertainty (i.e., imperfect management control). Probabilistic approaches have been proposed previously for setting target and limit reference points in fishery management. In this talk, we describe two adaptations to those earlier approaches designed for better consistency with recent revisions to the National Standards Guidelines. One adaptation is intended for setting ABC for a single year, the other for setting ABCs and ACTs over multiple years.

PSA. R. Wakeford, MRAG Americas, Inc.

No abstract available.

Reviewer's views

It was not clear to me what this ToR was asking. Risk evaluations methods were not explicitly reviewed, except the somewhat subjective risk assessment provided by PSA. The p* approach was reviewed and discussed. It seemed to be the favored method of incorporating uncertainty in management advice and recommendations.

The p* approach was clearly explained and seemed consistent with National Standards Guidelines.

PSA is a risk-based framework, primarily for data poor stocks, to evaluate species vulnerability. Risks are based on subjective scoring. I did not see any mechanism to deal with uncertainty. Presumably distributions of scores for the productivity and susceptibility attributes could be supplied for a subjective assessment of uncertainty; that is, distributions of overall scores could be generated (using monte-carlo resampling) based on the distributions of scores for the various attributes.

ToR 6: Prepare a SEDAR procedures document addressing these recommendations that will be used to guide future SEDAR assessments and projection analyses.

The document was not completed at the time of this review.

Summary of conclusions and recommendations

ToR 1: Review and discuss past approaches to characterizing and presenting uncertainty in stock assessments and projection analyses conducted under the SEDAR process.

Past approaches to characterizing and presenting uncertainty in stock assessments and projection analyses conducted under the SEDAR process were adequately reviewed. More attention should have been given to the deficiencies in these approaches, to see if there are common problems for the workshop to address.

Ad hoc approaches have been used to characterize uncertainty about stock status and projections under the SEDAR process. This is typical of most stock assessments, and simply reflects the different data available for assessments and the evolving nature of the theory and software technology available for quantifying uncertainty.

Sensitivity runs are often used to characterize uncertainty in the SEDAR process; however, sensitivity runs really address robustness and not uncertainty. In many situations the range of results from sensitivity analyses is not relevant or useful for advice; that is, one cannot conclude that there is a low probability that stock status is outside of this range.

If sensitivity analyses are provided to quantify uncertainty then the assessment workshop (AW) should also provide some quantification of the likelihood or plausibility (i.e. weightings) of each scenario. AW participants will often be the most informed group to do this.

Informative priors need to be rigorously challenged. Priors should not be based on the assessment model data because this is essentially 'using the data twice' which leads to a false measure of precision that is too high.

While over-simplified models may give false and too precise measures of uncertainty, overly complicated models may do the same. In highly parameterized models it is important to verify that confidence/credible intervals are not biased even if modeling assumptions are true.

Usually statistical inferences for current stock size will not enjoy the distributional robustness afforded by asymptotic arguments.

Including only recruitment uncertainty in projections may be insufficient, especially for short-term forecasts of fisheries that are not heavily dependent on recruitment (i.e. not a recruitment fishery). A difficult issue is the auto-correlation in recruitment residuals that is often apparent.

Additional process error in stock dynamics, and implementation error in regulating catches, is also important to consider when deriving benchmarks. If actual catches vary from the target or if there is process error in the model then deterministic MSY catches will not be sustainable in the long term.

ToR 2: Review and discuss alternative approaches to characterizing and presenting uncertainty in stock assessments and projection analyses, including those utilized in other regions. Discuss the sources of uncertainty which require consideration in the stock assessment process. Sources may include implementation uncertainty, within model uncertainty, inter-model uncertainties. Make recommendations on which sources of uncertainty to consider for future SEDAR stock assessments and projection analyses.

Alternative approaches to characterizing and presenting uncertainty in stock assessments and projection analyses, including those utilized in other regions, were partially reviewed.

A succinct summary of important uncertainties to consider in a stock assessment was not reviewed at the workshop; however, many sources of uncertainty were considered. The workshop did not specify a “must do” list of sources of uncertainty to consider for future SEDAR stock assessments and projection analyses.

ToR 3: Recommend approaches for representing uncertainty in the assessment documentation and expressing confidence in estimated parameters. Discuss both inter- and intra-model uncertainty. Include guidance for different model classes.

The theory and technology for representing uncertainty is still evolving which makes it difficult to recommend best practices, especially when the efficacies of the various approaches are not well understood.

The workshop did not make specific recommendations for expressing confidence in estimated parameters. Instead, a range of options were discussed. Also, a range of options was presented for accounting for inter- and intra-model uncertainty.

Instead of sensitivity analyses, use sensitivity distributions in models. Assume a distribution for uncertain inputs or parameters that fully reflects uncertainty about its value. If sensitivity analyses indicate the data are uninformative about a parameter (e.g. M) then do not try to update the sensitivity distribution (i.e. prior) using the data. If the data are informative then Bayesian or Empirical-Bayesian approaches should be used.

A lack of retrospective pattern does not mean there are no problems with model assumptions.

It is important to include within-survey variance in stock assessment models, but it is not straight-forward how to do this.

Continue development of the spreadsheet using a small number of case studies. An important goal should be to provide as much objective information as possible to assist in specifying prior distributions.

The empirical approach to quantifying stock assessment error assumes that the model specification error is constant over time and can be estimated using historic assessment results. There are reasons why this error may not be constant over time. This approach does not directly consider the validity of assessment assumptions (primarily reflected in fits to the data). The empirical approach requires further study.

Model averaging of sensitivity analyses to account for model uncertainty will be problematic because of uncertainty in subjective weights used for model averaging. Using model fits to assessment data to determine weights may also have problems, including robustness.

If the sources of model error are random variations in input parameters and data, then models with process error should, in aggregate, reasonably account for these uncertainties. If the mean of the errors is not zero then adding process error will probably not be sufficient.

ToR 4: Review and discuss uncertainty estimates needed for management, specifically addressing the needs for each council's ABC control rules and ACT determinations. Make recommendations on which uncertainty estimates should be included in future SEDAR stock assessments and projection analyses

The workshop reviewed and discussed uncertainty estimates needed for the South Atlantic Fishery Management Council and the Gulf of Mexico SSC ABC control rules and ACT determination.

The workshop did not make specific recommendations on which uncertainty estimates should be included in future SEDAR stock assessments and projection analyses.

ToR 5: Provide recommendations on best use of uncertainty characterization and recommendation methods such as P analysis, risk evaluation approaches, PSA*

Risk evaluations methods were not explicitly reviewed, except the somewhat subjective risk assessment provided by PSA. The p* approach was reviewed and discussed. It seemed to be the favored method of incorporating uncertainty in management advice and recommendations.

PSA did not account for uncertainty.

ToR 6: Prepare a SEDAR procedures document addressing these recommendations that will be used to guide future SEDAR assessments and projection analyses.

The document was not completed at the time of this review.

Critique of the NMFS review process

No comments

Appendix 1: Bibliography of materials provided for appointee's involvement

No documents were provided.

Appendix 2: A copy of the CIE Statement of Work

Attachment A: Statement of Work for Dr. Noel Cadigan

External Independent Peer Review by the Center for Independent Experts

SEDAR Procedural Workshop on Uncertainty

Scope of Work and CIE Process: The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Technical Representative (COTR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. The CIE reviewer is selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. The CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.com.

Project Description: SEDAR procedural workshops provide an opportunity for focused discussion and deliberation on topics that arise in multiple assessments and are structured to develop best practices for addressing common issues across assessments. The fourth procedural workshop will consider methods of addressing uncertainty in SEDAR assessments, including topics such as developing appropriate confidence intervals in both parameter estimates and projection outputs, methods of characterizing and expressing assessment uncertainty beyond that reflected in confidence intervals, use of sensitivity analyses and recommendations on standard sensitivities, and relating uncertainty to overall risk evaluation and especially the risk of overfishing occurring. The SEDAR Steering Committee considers this a critical topic and approved this workshop but planning was delayed due to a lack of funding in the 2009 planning budget. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

Requirements for CIE Reviewer: One CIE reviewer shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. The CIE reviewer shall have working knowledge and recent experience in the application of stock assessment, statistics, fisheries science, and marine biology sufficient to complete the primary task of participation in discussions of technical details of the methods used in characterizing and documenting uncertainty in SEDAR assessments. The CIE reviewer's duties shall not exceed a maximum of 14 days to complete all work tasks of the peer review described herein.

Location of Peer Review: The CIE reviewer shall conduct an independent peer review during the panel review meeting in Charlotte, North Carolina during 22-26 February 2010.

Statement of Tasks: The CIE reviewer shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

<u>Prior to the Peer Review</u>: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the COTR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and ToRs to the CIE reviewer. The NMFS Project Contact is responsible for providing the CIE reviewer with the background documents, report, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact is also responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COTR prior to the commencement of the peer review.

<u>Foreign National Security Clearance</u>: When the CIE reviewer participates during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for CIE reviewer who is a non-US citizens. For this reason, the CIE reviewer shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/sponsor.html).

<u>Pre-review Background Documents</u>: Two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the CIE reviewer the necessary background information and report for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE Lead Coordinator on where to send documents. The CIE reviewer is responsible only for the pre-review documents that are delivered to the reviewer in accordance to the

SoW scheduled deadlines specified herein. The CIE reviewer shall read all documents in preparation for the peer review.

Panel Review Meeting: The CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. **The CIE reviewer serves only as a peer reviewer in accordance with the SoW, and shall not serve as an analyst during the workshop. Modifications to the SoW and ToRs can not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COTR and CIE Lead Coordinator.** The CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewer as specified herein. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

Contract Deliverables - Independent CIE Peer Review Report: The CIE reviewer shall complete an independent peer review report in accordance with the SoW. The CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. The CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

Other Tasks – Contribution to Summary Report: The CIE reviewer may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. The CIE reviewer is not required to reach a consensus, and should provide a brief summary of the reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Specific Tasks for CIE Reviewer: The following chronological list of tasks shall be completed by the CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and report provided by the NMFS Project Contact in advance of the peer review.
2) Participate during the panel meeting in Charlotte, North Carolina during 22-26 February 2010.
3) In Charlotte, North Carolina during 22-26 February 2010 as specified herein, and conduct an independent peer review in accordance with the ToRs (**Annex 2**).
4) No later than 11 March 2010, the CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and David Sampson, CIE Regional Coordinator via email to david.sampson@oregonstate.edu. The CIE report shall be written using the

format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

19 January 2010	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
9 February 2010	NMFS Project Contact sends the CIE Reviewer the pre-review documents
22-26 February 2010	The reviewer participates and conducts an independent peer review during the panel review meeting
12 March 2010	CIE reviewer submits draft CIE independent peer review report to the CIE Lead Coordinator and CIE Regional Coordinator
26 March 2010	CIE submit CIE independent peer review report to the COTR
02 April 2010	The COTR distributes the final CIE report to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: Requests to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the COTR within 10 working days after receipt of all required information of the decision on substitutions. The COTR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewer to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: Upon review and acceptance of the CIE independent peer review report by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these report shall be sent to the COTR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review report) to the COTR (William Michaels, via William.Michaels@noaa.gov).

Applicable Performance Standards: The contract is successfully completed when the COTR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:

(1) The CIE report shall completed with the format and content in accordance with **Annex 1**,
(2) The CIE report shall address each ToR as specified in **Annex 2**,

(3) The CIE report shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Distribution of Approved Deliverables: Upon acceptance by the COTR, the CIE Lead Coordinator shall send via e-mail the final CIE report in *.PDF format to the COTR. The COTR will distribute the CIE report to the NMFS Project Contact and Center Director.

Key Personnel:

William Michaels, Contracting Officer's Technical Representative (COTR)
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-713-2363 ext 136

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

Julie Neer, SEDAR 22 Coordinator, NMFS Project Contact
South Atlantic Fishery Management Council
4055 Faber Place, Suite 201
North Charleston, SC 29405
Julie.Neer@SAFMC.net Phone: 843-571-4366

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

   a. Reviewer should describe in their own words the review activities completed during the panel review meeting, including providing a brief summary of findings, of the science, conclusions, and recommendations.

   b. Reviewer should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views.

   c. Reviewer should elaborate on any points raised in the Summary Report that they feel might require further clarification.

   d. Reviewer shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

   e. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The CIE independent report shall be an independent peer review of each ToRs, and shall not simply repeat the contents of the summary report.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE Statement of Work
   Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

SEDAR Procedural Workshop on Uncertainty

Review and discuss past approaches to characterizing and presenting uncertainty in stock assessments and projection analyses conducted under the SEDAR process.

Review and discuss alternative approaches to characterizing and presenting uncertainty in stock assessments and projection analyses, including those utilized in other regions. Discuss the sources of uncertainty which require consideration in the stock assessment process. Sources may include implementation uncertainty, within model uncertainty, inter-model uncertainties. Make recommendations on which sources of uncertainty to consider for future SEDAR stock assessments and projection analyses.

Recommend approaches for representing uncertainty in the assessment documentation and expressing confidence in estimated parameters. Discuss both inter- and intra-model uncertainty. Include guidance for different model classes.

Review and discuss uncertainty estimates needed for management, specifically addressing the needs for each council's ABC control rules and ACT determinations. Make recommendations on which uncertainty estimates should be included in future SEDAR stock assessments and projection analyses.

Provide recommendations on best use of uncertainty characterization and recommendation methods such as P* analysis, risk evaluation approaches, PSA.

Prepare a SEDAR procedures document addressing these recommendations that will be used to guide future SEDAR assessments and projection analyses.

Annex 3: Tentative Agenda

SEDAR Procedural Workshop on Uncertainty

Charlotte, North Carolina during 22-26 February 2010

Monday, February 22, 2010:

1:00 – 3:00pm	Intro and presentations
3:00-3:30pm	Afternoon Break
3:30 – 6:00pm	Presentations

Tuesday, February 23, 2010:

8:30 – 10:00am	Presentations
10:00-10:30am	Morning Break
10:30- 12:00pm	Discussions
12:00 – 1:30pm	Lunch Break
1:30 - 3:30pm	Presentations
3:30 – 4:00pm	Afternoon Break
4:00 – 6:00pm	Discussions

Wednesday, February 24, 2010:

8:30 – 10:00am	Presentations
10:00-10:30am	Morning Break
10:30- 12:00pm	Discussions
12:00 – 1:30pm	Lunch Break
1:30 - 3:30pm	Discussions
3:30 – 4:00pm	Afternoon Break
4:00 – 6:00pm	Discussions

Thursday, February 25, 2010:

8:30 – 10:00am	Presentations
10:00-10:30am	Morning Break
10:30- 12:00pm	Discussions
12:00 – 1:30pm	Lunch Break
1:30 - 3:30pm	Discussions
3:30 – 4:00pm	Afternoon Break
4:00 – 6:00pm	Discussions

Friday, February 26, 2010:

8:30 – 10:00am	Report Writing
10:00-10:30am	Morning Break
10:30- 1:00pm	Report Writing

Appendix 3: Membership or other pertinent information from the panel review meeting.

References

Bousquet, N., Duchesne, T. and Rivest, L.-P. 2008. Redefining the maximum sustainable yield for the Schaefer population model including multiplicative environmental noise. J. Theor. Biol. 254 65-75.

Bousquet, N., Cadigan, N., Duchesne, T., and Rivest, L.-P. 2010. Detecting and correcting underreported catches in fish stock assessment. Revised for Can. J. Fish. Aquat. Sci.

Cadigan, N. G. and Farrell, P. J. 2002. Generalized local influence with applications to fish stock cohort analysis. Appl. Statist. 51: 1-15.

Fahrmeir, L., and Kaufmann. H. 1985. Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models. The Annals of Statistics, Vol. 13, No. 1, pp. 342-368.

Punt, A.E., and Hilborn, R. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. Rev. Fish Biol. Fish. 7: 35-63.

Patterson, K., Cook, R., Darby, C., Gavaris, S., Kell, L., Lewy, P., Mesnil, B., Punt, A., Restrepo, V., Skagen D.W., and Stefansson, G. 2001. Estimating uncertainty in fish stock assessment and forecasting. Fish Fish. 2: 125-157.

Severini T.A. 2000. Likelihood Methods in Statistics, Oxford University Press, Oxford.

Appendix 4: Reviewer's summary of sources of uncertainty in stock assessment

This review is mostly derived from Patterson et al. (2001), but I included some additional items. It does not include implementation error in management advice.

1. structural models
    - mortality
        - natural: distribution of values, and is it age and time invariant?
        - Fishing: Are there errors in landings and their age-compositions? Is fishing mortality separable, or separable within gear-types? or separable with process error? Do these errors vary smoothly in time or age? Shrinkage or

parameter penalties may introduce undesirable biases both in the parameters and in the variance estimates.

- Catchability of stock size indices. Is it time invariant but age-specific? Is it dome-shaped or flat-topped? Does it depend on stock size (esp. for CPUE)? Are there year-effects in catchability?
- growth and maturation. Are these processes time invariant? Is there density dependence? Are there environmental effects?
- dependence of recruitment on stock size. Does it exist or is recruitment unpredictably episodic? What is the form of the relationship? Is there depensation? Extrapolation errors? Regime shifts?
- ecosystem considerations. Predation? Regime shifts?

2. Error models
   - Total catch. Bias (mis-reporting, discards)? Precision? Are these factors time-invariant? Management effects?
   - Age-composition. Distribution (multinomial, Dirichlet-multinomial, Gaussian-multinomial etc)? Effective sample size due to cluster sampling?
   - Effort. Fishing power? Hours fished? mixed-species fisheries? Gear? Skill? Standardization? Spatial and temporal resolution?
   - Indices. Distribution? Zero's?
   - Variance for different data sources. Internal or external estimates? Components of variance (measurement and model error)? Time-varying? Age-vary? Index coverage?
   - Weight and maturity at age. Modeled or not? Sampling error? modeling error?
   - Recruitment. How to resample residuals? Auto-correlation?
3. Inference paradigm and estimation methods
   - Bayesian. Priors? Robust? Posterior calculation (MCMC convergence)? Accuracy of inferences?
   - Frequentist. Robust? Confidence distribution calculation (delta method, Profile likelihood, Bootstrap method)? Accuracy of inferences?
   - Observation and process error?

Patterson et al. (2001) concluded that uncertainty estimation techniques should meet four criteria:

1. The method should be properly conditioned, i.e. it should only take as axiomatic assumptions which either have a very low probability of being wrong or for which any mis-specification has very little consequence
2. The distributional assumptions made about parameters or data admitted as uncertain should be validated.
3. If there remain significant uncertainties due to alternative model structures or alternative conditioning choices being perceived as having high probability, then such uncertainties should be transmitted clearly.

4. The conditioning decisions should dictate the estimation method, rather than institutional convention or use of standard software.

In practice it is often the case that a fairly wide range of structural models, often with very different consequences for management, may appear to represent the data almost equally well, and an ideal of proper conditioning is not often attainable. In such cases of indeterminacy, it is most important that uncertainty due to model structure should be presented in an interpretable form. Possible Approaches: Model averaging. Decision Tables.